MARIA HEIKKILÄ

SYSTEMATIZING MORPHOLOGY:

A TOTAL EVIDENCE APPROACH TO DITRYSIAN PHYLOGENETICS (LEPIDOPTERA)

FINNISH MUSEUM OF NATURAL HISTORY
FACULTY OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES
UNIVERSITY OF HELSINKI

Systematizing morphology:

a total evidence approach to ditrysian phylogenetics (Lepidoptera)

Maria Heikkilä

Finnish Museum of Natural History
Zoology Unit
University of Helsinki
Finland

Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Finland

Academic Dissertation

To be presented for public examination with the permission of the
Faculty of Biological and Environmental Sciences,
University of Helsinki,
in the Auditorium C205, Economicum (Arkadiankatu 7),
on December 19th 2014 at 12 noon.

Helsinki 2014

Supervised by:

Senior Curator, Dr. Lauri Kaila,
Finnish Museum of Natural History, Zoology Unit,
University of Helsinki, Finland.

Reviewed by:

Dr. Joaquin Baixeras,
University of Valencia,
Instituto Cavanilles Biodiversidad y Biología Evolutiva, Spain.

Prof. Toomas Tammaru,
University of Tartu, Estonia.

Examined by:

Dr. Jurate De Prins,
Natural History Museum London, U.K.

Custos:

Prof. Jouni Laakso
University of Helsinki, Finland

Author's address:
Finnish Museum of Natural History, Zoology Unit
P.O. Box 17, FI-00014, University of Helsinki,
Finland
maria.heikkila@helsinki.fi

Cover design: Kaisaleena Halinen
Layout: Heidi Viljanen

ISBN 978-951-51-0495-3 (paperback)
ISBN 978-951-51-0496-0 (PDF)
http://ethesis.helsinki.fi
Painosalama Oy
Turku 2014

My work enraptures but utterly exhausts me… To know that no one before you has seen an organ you are examining, to trace relationships that have occurred to no one before, to immerse yourself in the wondrous crystalline world of the microscope, where silence reigns, circumscribed by its own horizon, a blindingly white arena – all this is so enticing that I cannot describe it.

Excerpt from a letter from Vladimir Nabokov (novelist and lepidopterist) to his sister, Elena Sikorski, November 25, 1945.
In Selected Letters, p. 58–59.

Contents

List of original articles

This thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I. **Heikkilä, M.** and Kaila, L. 2010. Reassessment of the enigmatic lepidopteran family Lypusidae (Lepidoptera: Tineoidea; Gelechioidea). Systematic Entomology. 35 (1): 71-89.

II. **Heikkilä, M.**, Kaila, L. Mutanen, M., Peña, C. and Wahlberg, N. 2012. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. Proceedings of the Royal Society B. Biological Sciences. 279: 1093-1099.

III. **Heikkilä, M.**, Mutanen, M., Kekkonen, M. and Kaila, L. 2014. Morphology reinforces proposed molecular phylogenetic affinities: a revised classification for Gelechioidea (Lepidoptera). Cladistics – the international journal of the Willi Hennig Society. 30 (4): 563-589.

IV. **Heikkilä, M.**, Mutanen, M., Wahlberg, N., Sihvonen, P., and Kaila, L. Pinning down rogues? Combining morphological and molecular data in ditrysian Lepidoptera. (manuscript).

Table of contributions

The following table indicates the major contributions of authors to the original articles and the manuscript.

	I	II	III	IV
Original idea	LK	MM	LK, MM	LK, PS
Study design	LK	**MH**, NW, CP, LK, MM	LK, **MH**	LK, PS, **MH**, MM, NW
Morphological data	**MH**, LK	**MH**, LK	LK, **MH**, MK	**MH**, LK, PS
Molecular data	-	MM, NW	MM	MM, NW
Analyses	**MH**, LK	NW, CP, **MH**, MM	**MH**, LK	**MH**, LK, NW, MM
Manuscript preparation	**MH**	**MH**, NW, LK, CP, MM	**MH**, LK, MM, MK	**MH**, LK, MM, NW, PS

MH=Maria Heikkilä, LK= Lauri Kaila, MK= Mari Kekkonen, MM= Marko Mutanen, CP=Carlos Peña, PS= Pasi Sihvonen, NW= Niklas Wahlberg

Disclaimer
This publication is not available for purposes of zoological nomenclature in accordance with the International Code of Zoological Nomenclature, fourth edition, article 8.3.

Abstract

Lepidoptera (butterflies and moths) are one of the most diverse and species-rich groups of organisms. Currently, the number of described species is close to 160.000, but it is estimated that the true number of species could be as high as half a million (van Nieukerken *et al.* 2011). Nearly 99% of all Lepidoptera belong to the more advanced subclade Ditrysia (Kristensen & Skalski 1999). The focus of this thesis is the evolutionary history of the enormous ditrysian clade. The relatedness between the 30 ditrysian superfamilies has been a mystery that has only recently begun to unfold. The aim of the thesis is to find new information on the evolutionary relationships between these superfamilies, and within two of the largest of them, Gelechioidea and Papilionoidea, both with over 18.000 described species. Evidence on the evolutionary relationships is sought from both morphological and genetic data. The morphological datasets are based on characters coded from larval, pupal and adult stages, and are thus far the largest in terms of the number of coded characters and of the number of exemplar species.

The phylogenetic analysis in chapter I is based solely on morphological character data. In the analyses of chapters II-IV, morphological data are combined to molecular data from eight gene regions sequenced from corresponding species or genera. Morphological characters supporting phylogenetic affinities are discussed in detail. Characters described in the articles of this thesis can be used in the identification of extant and fossil taxa, and in future studies focusing on the evolution of specific morphological traits.

Chapter I elucidates the evolutionary affinities of the mysterious genus *Lypusa* (Lypusidae), the phylogenetic position of which has not been known. It is demonstrated that this genus actually belongs within the megadiverse superfamily Gelechioidea.

Chapter II focuses on the phylogenetic relationships between the families of the superfamily Papilionoidea (Papilionidae, Pieridae, Nymphalidae, Lycaenidae, Riodinidae, Hesperiidae and Hedylidae). The results show that the traditional concept of butterflies, i.e. including only the first five above-mentioned families, is paraphyletic. Hesperiidae and Hedylidae are sister-groups and are internested within the rest of the butterflies. A divergence time analysis indicates that this superfamily originated in the Cretaceous and diversified in the Tertiary.

Chapter III explores the evolutionary relationships within one of the largest but least studied groups of Lepidoptera, Gelechioidea. The classification of this superfamily has been known to be in the need of revision. Based on the results, a new classification into 16 families is proposed.

The aim of chapter IV is to find morphological evidence on the relationships between the ditrysian superfamilies. Several morphological characters claimed to define subgroups of Ditrysia are tested. New characters supporting evolutionary affinities of clades are described and discussed. By combining morphological and sequence data, some taxa that have been unstable in analyses based on either type of data only, find a stable position, e.g. Epipyropidae and Cyclotornidae, Urodoidea + Schreckensteinioidea.

Summary

1. Introduction

1.1 Phylogenetic systematics

The field of biology that studies the diversification and relatedness of organisms is called systematics. Systematics in general resembles historical rather than empirical studies. This is because we cannot study the past diversification events themselves, but have only clues left to reconstruct these events, some of which have happened millions of years ago. Such clues are the fragmentary fossil record, but also the morphological, molecular, behavioral and ecological characters that extant species have inherited from their ancestors (Wenzel 2002). To explain how organisms are related to each other, we must make inferences about their relationships by studying these characters and find traces revealing descent from a common ancestor. Organisms sharing homologous characters of the same kind are likely to share a common ancestor and are thus closely related.

When the evolutionary relatedness between organisms of interest is studied, such clues are collected from representative organisms. The molecular, morphological and behavioral characters serve as potential evidence for grouping and inferring evolutionary relationships between them. The collected data, usually arranged as matrices, are analyzed using various computing methods. As the number of organisms in the study and the amount of data grow, inferring a phylogeny becomes more and more difficult and computationally demanding. The analyses yield as outcome a branching diagram, a phylogenetic tree. Phylogenetic trees, or phylogenies, represent the evolutionary relationships between organisms. The organisms in the study are at the tips of the branches and the points where the branches meet, the nodes, represent the hypothetical common ancestors. Descendants that split from the same node are more closely related to each other than to any other organism in the study. Phylogenies are thus hypotheses on the evolutionary history of organisms based on the evidence available. They can be seen as chronicles of evolutionary events that have led to the current diversity and distribution of organisms (O'Hara 1988).

Phylogenies are used as the foundation for classifying monophyletic groups (i.e. groups including a common ancestor and all its descendants) and describing taxonomic levels (species, genus, tribe, family, superfamily, etc.). Well-founded phylogenetic trees serve also as the analytical basis for research on the evolutionary history of organisms and for various studies with evolutionary aspects, such as the evolution of specific traits (morphological, physiological, behavioral or ecological). Knowledge of the phylogeny is thus a necessary component for understanding the behavior, ecology and evolution of single species.

Phylogenetic information is also being used in conservation planning aiming

at preserving phylogenetic diversity instead of just species diversity, and in predicting extinction risk across different phylogenetic lineages. Some lineages may be more vulnerable to extinction than others if the attributes making them susceptible are shared due to a common evolutionary history (Rolland *et al.* 2012).

1.2 Morphology and genes

For centuries, humans have attempted to achieve a better understanding of the outstanding diversity of life by arranging organisms into groups according to their structural similarities and differences. Comparative morphology was of central role also in the works developed and formalized especially by Carl von Linné (1707-1778) aiming at classifying organisms into hierarchical groups. It was also the observations on similarities and differences between organisms that formed the basis for the theory of evolution and common descent of organisms by Charles Darwin (1809-1882) and Alfred Russel Wallace (1823-1913). Comparative morphology has historically been, and still largely continues to be, the basis in the description of species, yielding an uncounted amount of valuable literature and illustrations.

For decades, the physical characteristics of organisms provided almost the only source of data for the reconstruction of evolutionary trees. In recent years, however, methods exploring DNA have revolutionized research on the history of life. Large quantities of information on the relatedness of organisms present in the genes have become easily accessible. The new information genetic material has shed on the relationships of organisms is uncontested. Molecular datasets, sometimes consisting of the whole genome (i.e. millions of characters) of the organisms under study, definitely outweigh in size the laboriously created morphological datasets, which at most reach some hundreds of characters. It is, however, less clear to what extent the relative information content of morphological and molecular data can be compared.

Although new techniques allow the acquisition of larger quantities of morphological and anatomical data, it is evident that with the increasing ease and feasibility of obtaining molecular datasets the role of comparative morphology, being more laboursome and requiring long training, has been dwindling in studies of evolutionary biology (de Carvalho *et al.* 2008; Giribet 2010). Nevertheless, understanding of morphology remains crucially important. New and original research on the morphology of organisms is essential: if morphological data are only taken from earlier work, existing hypotheses remain untested and new diagnostic characters may be left undiscovered (Wheeler 2008).

Morphological characters are also the only bridge between fossil and extant species, and understanding the evolution of specific structural traits allows us to assign fossils to a certain evolutionary lineage (Wiens 2004; Hermsen & Hendricks 2008). Identified and dated fossils may be used

as calibration points in obtaining date estimates for the evolutionary tree and the branching events within it. A dated phylogeny enables the exploration of effects of the geological, climatic or biotic history of the Earth on the evolution of organisms (Drummond *et al.* 2006; Forest 2009).

It is nowadays widely accepted that different data should be combined in phylogenetic analyses as they contribute to different nodes in the phylogeny (Rokas *et al.* 2003). In total-evidence studies, where all the data available are combined, a weak phylogenetic signal may become stronger and thus expose a "secondary signal" (Nixon & Carpenter 1996) or "hidden support" (Gatesy 2005). In some cases, morphological characters have also been shown to be able to anchor species in a phylogeny that have not found a stable position in analyses based on molecular data only (Heikkilä *et al.* 2014). Morphological data can thus contribute to the phylogenetic analysis. Nowadays appropriate models and corrections exist to incorporate morphological data to model-based analyses used to analyze genetic data (Lewis 2001; Wright & Hillis 2014). Often morphological or behavioral characters are only mapped onto the tree based on DNA or other sequence data serving as a scaffold to study the evolution of certain traits. In such cases the information content the morphological data has to offer in the reconstruction of the evolutionary tree is not taken full advantage of (Hermsen & Hendricks 2008). Congruent morphological and genetic evidence of course strengthen the reliability of a phylogenetic hypothesis. Likewise, conflicting evidence encourages looking for explanations, verifying and seeking for more data to test the hypotheses.

1.3 Lepidoptera: butterflies and moths

Half of all named organisms are insects (Scudder 2009) and one of the four most diverse orders in this class are Lepidoptera (moths and butterflies) (Kristensen *et al.* 2007). Currently, the number of described species of Lepidoptera is close to 160.000 but it is estimated that the actual number of species could climb close to 500.000 (van Nieukerken *et al.* 2011). Lepidoptera belong to the holometabolous insects, meaning that they undergo complete metamorphosis during their development. The adult moths and butterflies are characterized by the scale cover of the wings and body, which inspired Linné to give them the name Lepidoptera (derived from Greek, lepis, lepidos = scale, pteron = wing). The order Lepidoptera has a worldwide distribution. The lineages considered to be the oldest seem to have a trend of being most diverse in the temperate regions of the world (although this could be due to tropical undersampling (Lopez-Vaamonde *et al.* 2014)). In more advanced lineages, however, diversity is highest in the tropics (Holloway & Nielsen 1999).

Like other insects, Lepidoptera are a group of organisms that carry out important biotic interactions and in that way are a necessary factor for ecological functioning. Some lepidopteran larvae feed on detritus

and fungi, and some are even predatory, but by far most of their larvae consume living plant material. Together with the beetle clade Phytophaga, Lepidoptera form the largest radiation of plant feeding insects, making their ecological impact considerable (Scoble 1992; Powell *et al.* 1999). Larvae, pupae and adult Lepidoptera form also an important source of food for animals such as birds, bats, other small mammals, some insects, numerous parasitoids and even humans. Actually, it has been estimated that globally more people use insects as food than dairy products (DeFoliart 1989; Scoble 1992; Ramos-Elorduy *et al.* 2011).

Apart from being the object of admiration and fervent collecting, Lepidoptera are present in the lives of humans in many other ways. A great number of lepidopteran species with herbivorous larvae pose a threat as serious pests of crops (e.g. the Diamondback moth, *Plutella xylostella* on vegetables (Talekar 1993)) and forests (e.g. *Lymantria monacha* in Central Europe (Kitching & Rawlins 1999)). On the other hand, they are also used in biological control to contain or eradicate invasive plant species. Although most Lepidoptera lack specialized organs for pollen collecting, they have been ranked the fourth most important group of pollinators (Barth 1985), an aspect that can be counted as an ecosystem service (Winfree *et al.* 2011).

Lepidoptera have also been exploited in silk production for centuries, in particular the silkworm moth *Bombyx mori* (Scoble 1992). In addition, the silkworm moth has had medicinal importance in traditional Chinese medicine, and the utility of the medicinal substances related to these moths are being re-explored along with those of other species (Chiu 2003; Kikuchi *et al.* 2004). Lepidoptera are also used as model organisms in many fields of science, e.g. in ecological modeling, in developmental and genetic studies (Roe *et al.* 2010) and serve as environmental indicators (e.g. Fox *et al.* 2011). The full potential of the results of such studies can, however, be obtained only when the evolutionary history of the Lepidoptera is better resolved and a phylogenetic framework to extrapolate the results across larger groups becomes possible.

1.4 Ditrysian Lepidoptera and research on their evolutionary history

The oldest lepidopteran fossil is from the Jurassic period, and thus estimated to be about 190 million years old (Whalley 1985; Grimaldi 1999; Grimaldi & Engel 2005a). Divergence time analyses push the origin of Lepidoptera to about 230 million years ago (Wahlberg *et al.* 2013). These estimates make Lepidoptera a relatively recently radiated order when compared to the advent of insects, which is thought to have happened 400 million years ago (Grimaldi & Engel 2005b). For comparison, tetrapods (the four-limbed vertebrates) originated about 350 million years ago.

Around 150 Million years ago, the ditrysian clade, which is the lineage comprising close to 99% of all butterflies and moths

living today, diverged from the more basal Lepidoptera, and began to diversify (Kristensen & Skalski 1999; Wahlberg *et al.* 2013). The monophyly of Ditrysia is supported by the name giving character (di- = two, trūpa, trus- = a hole (Emmet 1991)) expressed through the anatomy of the females, which have two genital orifices, one for mating and one for laying eggs as opposed to the females of the more basal monotrysian Lepidoptera, which only have one aperture (Dugdale 1974).

The phylogenetic relationships between species and genera belonging to the Ditrysia have been much studied by both amateur and professional lepidopterists. Comparative morphology has held an important role in species description and systematics, but in recent years better computing methods have also enabled larger-scale phylogenetic studies. These studies have helped gain insight into the evolutionary relatedness between families within several of the larger ditrysian superfamilies (e.g. Kaila 2004; Bucheli & Wenzel 2005; Niehuis *et al.* 2006; Solis 2007; Kawahara *et al.* 2008; Regier *et al.* 2008; Warren *et al.* 2009; Kaila *et al.* 2011; Zahiri *et al.* 2011; Heikkilä *et al.* 2012; Heikkilä *et al.* 2014; Regier *et al.* 2014).

The relationships at the higher level, between the ditrysian superfamilies, have also been the object of much research in recent years. The division of Ditrysia into superfamilies by morphological characters has remained relatively stable and more or less generally agreed upon. In van Nieukerken *et al.* (2011), Ditrysia are divided into about 30 superfamilies, but several of these are monotypic (meaning that they include only one species, genus or family) and have been assigned a superfamily of their own because their relation to other Lepidoptera is not known. Also, a number of enigmatic genera or families have eluded a well-supported affiliation to any of the recognized superfamilies and are currently without a designated superfamily.

The difficulties in resolving the relationships within the ditrysian Lepidoptera have been attributed to the homogeneous morphology of Ditrysia and the lack of obvious morphological characters, which could reveal the relatedness of the superfamilies (Kristensen & Skalski 1999). In 1999 Kristensen and Skalski published a very tentative phylogeny for the ditrysian Lepidoptera based on the current knowledge, notably the works in comparative morphology by Minet (1986, 1991). In this phylogeny only very few relationships between the superfamilies could be proposed, the division of the clade into three inter-nested subclades, Apoditrysia, Obtectomera and Macrolepidoptera being the major ones. To identify and place a lepidopteran to a level above the species level requires often exposing the exoskeleton that is hidden by scales. This laboriousness combined with the large number of species featuring different combinations and unclear limits of such characters have hindered their classification (Kristensen *et al.* 2007).

Since the tentative phylogeny presented by Kristensen and Skalski several large-scale molecular studies based on either nuclear

or mitochondrial DNA, or transcriptomic data, have addressed the problem of solving the evolutionary chronicle of the ditrysian lineages (Regier *et al.* 2009; Mutanen *et al.* 2010; Cho *et al.* 2011; Bazinet *et al.* 2013; Regier *et al.* 2013; Kawahara & Breinholt 2014; Timmermans *et al.* 2014). These studies have offered new hypotheses on the relationships between the superfamilies, some of them unexpected and contradicting current the composition and circumscription of superfamilies, and have revealed which parts of the phylogeny should be focused on more. Although the general patterns of the results of these studies are often in agreement, many of the proposed relationships still stand on uncertain grounds due to low support values or low taxon sampling.

A possible explanation for the initial difficulties in inferring a phylogeny for the Ditrysia, and several subgroups of it, is ancient rapid radiation (Whitfield & Kjer 2008). When lineage splitting within a relatively short time span in the ancient past has occurred, there has been little opportunity for molecular or morphological changes to build up. This meagre phylogenetic evidence of a close relationship between lineages can have been obscured or entirely vanished by subsequent changes that have accumulated since, inadequate data, or conflict with datasets (Rokas & Carroll 2006; Whitfield & Lockhart 2007). The radiation of ditrysian Lepidoptera in the mid to late Cretaceous into the Tertiary, a period coinciding with the great diversification of flowering plants (angiosperms), seems to have been relatively rapid. The phylogenetic signal may also have been muddled by extinction events, which have wiped out many lineages and phylogenetically informative characters.

Another reason for the difficulty to infer a stable phylogeny for Lepidoptera is their outstanding diversity. It has been impossible to include representatives of all groups in the analyses, especially in the very expensive phylogenomic studies, making by necessity some of the included taxa very distantly related, and others, perhaps crucial ones, omitted. Insufficient taxon sampling has been shown to be a cause of the so-called long-branch attraction, where distant taxa can be drawn together (Bergsten 2005; Kolaczkowski 2009; Talavera & Vila 2011, Boussau *et al.* 2014).

The effort to better understand the evolutionary history of the Ditrysia as a whole, and of the very diverse ditrysian superfamilies, is still an ongoing journey. To reach the destination, a more inclusive sampling of representative species and genes, combined with extensive analysis of morphological characters uniformly studied across the groups is essential. The key innovations that have enabled the success of this megadiverse group of insects are often expressed as morphological features. Mapping the distribution of structural character states, of the 'enabling mechanisms', allows to trace underlying evolutionary pathways, and as Kristensen (2011) puts it: "these are amongst the most rewarding uses that can be made of robust cladograms, irrespective of whether these were in the first place

generated from molecules, morphology or both".

2. Outline and aims of the thesis

The aim of this thesis is to bring new information on the phylogenetic affinities between the ditrysian superfamilies (Chapter IV) and selected groups within the clade; Gelechioidea (Chapters I & III) and Papilionoidea (Chapter II), and seek morphological characters to support these affinities and serve as diagnostic characters in the identification of taxa and groups of taxa.

The specific objective of Chapter I is to find a phylogenetic position for the enigmatic family, Lypusidae, the affinities of which had for long been unclear. The sole genus of this family, *Lypusa* Zeller, 1982 has been placed in several families within the superfamily Tineoidea, or a family of its own without known affinities. This study is entirely based on morphological data and reveals the phylogenetic position *Lypusa* within the huge and diverse superfamily Gelechioidea. The cladistic analysis places *Lypusa* with the Amphisbatidae. The morphological characters supporting this position are discussed.

Chapter II treats the phylogenetic affinities of butterflies in the traditional sense (Papilionidae, Pieridae, Nymphalidae, Lycaenidae and Riodinidae), the skippers, (Hesperiidae) and the New World butterfly-moths (Hedylidae). With over 18.000 described species, these families form one of the four largest superfamilies of Lepidoptera. The study is based on both DNA-sequence data and morphological data. The results support a sister-group relationship between the skippers and butterfly-moths, and their phylogenetic position within the butterflies, making Papilionoidea in the traditional sense paraphyletic. Using fossils as calibration points the divergence times of the lineages are estimated. This information, and knowledge on the current diversity of each of the families, are used to explore the tempo of diversification.

The aim of Chapter III is to perform a phylogenetic analysis and present a revised classification for the massive radiation of gelechioid moths, which currently includes over 18.000 described species. The superfamily Gelechioidea may well be the most species-rich group of Lepidoptera and its classification has been in great need of revision. The phylogenetic analyses (parsimony, maximum likelihood, Bayesian inference) are based on the largest dataset on the superfamily to date and include both morphological and sequence data. Based on the results of the study, a division of the superfamily into 16 families is proposed.

The objective of Chapter IV is to study the evolutionary relatedness of the 33 ditrysian superfamilies, which have for long been unresolved. The Ditrysia comprise nearly 99 % of all butterflies and moths and include over 160.000 described species. By combining a large morphological dataset and molecular data from eight gene regions, the relationships

between the superfamilies are inferred and compared to the results of recent studies based on genetic data. The aim is also to test the validity of several morphological characters previously proposed to define subclades of Ditrysia and search for more morphological evidence on the common ancestry of superfamilies. Problems arising from combining morphological and molecular data are discussed in the context of Lepidoptera.

3. Material and Methods

3.1 Material

The specimens used in the morphological studies (chapters I-IV) were obtained from several museums, private collections, and several persons, in particular Lauri Kaila and Marko Mutanen, were involved in collecting material (see Acknowledgements). For all exemplar species we tried to obtain all life stages; larva, pupa and adult male and female. The larvae were either dry-inflated or kept in alcohol. The pupae were dry or kept in alcohol. Sometimes the larval skin or pupal exuvia were used. The adult specimens were mounted.

The material for DNA extraction was also mostly collected by Lauri Kaila and Marko Mutanen or acquired from an international network of colleagues (see Acknowledgements).

3.2 Examination of morphology

The morphological examination of specimens was done with light microscopes; Leica MZ 75 stereomicroscope, magnification up to 400×, Wild M10 stereomicroscope, magnification up to 512×, Olympus SZX16 magnification up to 110×, and Leitz Diaplan phase contrast microscope (maximum magnification 1560×).

The larvae and pupae were examined for external characters. The examination of adult specimens was first done for external characters, such as scale vestiture. The wings of the specimens were then removed and the body treated in 10% potassium hydroxide (KOH) solution to dissolve protein and lipids, leaving the exoskeleton for examination. For the examination of wing venation, small wings were mounted on slides and large wings were wetted with drops of alcohol and scales brushed off to expose the veins. In some cases, slides were also made of other body parts, e.g. the abdomen. Morphological data were collected in matrices using WinClada (Nixon 2002) or MorphoBank (O'Leary & Kaufman 2012).

3.3 DNA extraction and sequencing

For the molecular analyses in Chapters (II-IV) one mitochondrial (cytochrome oxidase subunit I, COI) and seven protein-coding nuclear gene regions (Elongation factor-1a, EF-1a; Wingless; Ribosomal protein S5, RpS5; Cytosolic malate dehydrogenase, MDH; Glyceraldehyde-3-phosphate dehydrogenase GAPDH;

Carbamoyl phosphate synthetase domain protein CAD and Isocitrate dehydrogenase, IDH) were sequenced, totaling in over 6000 base pairs. DNA was extracted mostly from legs detached from the exemplar taxa, but sometimes from other body parts.

DNA amplification and sequencing followed the protocol presented in Wahlberg & Wheat (2008). Sequencing was performed with ABI 3500 and 3730 capillary sequencers. The VoSeq program was used to construct taxon sets (Peña & Malm 2012).

3.4 Phylogenetic analyses

To infer phylogenetic relationships both molecular and morphological data were analyzed with methods based on different optimality criteria and philosophies; model-based methods (maximum likelihood and Bayesian inference) and parsimony (except in Chapter I, where only parsimony was used). The use of different methods allows exploring how the data behaves and whether these methods find a similar phylogenetic signal.

Parsimony analyses were conducted with the program Tree search using New Technology, TNT, versions 1.0 and 1.1 (Goloboff *et al.* 2000, 2008), NONA (Goloboff 1993) and WinClada (Nixon 2002). Maximum likelihood analyses were performed with the program RAxML version 7.3.1 (Stamatakis 2006) and Bayesian inference with MrBAYES version 3.1 (Huelsenbeck & Ronquist 2001).

Models for the Bayesian analyses were obtained with FINDMODEL (Chapter II) (Tao et al. 2010) or PartitionFinder (Chapter III & IV) (Lanfear *et al.* 2012). In the maximum likelihood analyses, molecular data were analyzed under the GTR + Γ model. Data of the morphological partition were assigned the Mk model (Lewis 2001).

Several partition strategies were used in the model-based analyses. In chapters II-IV, molecular data were first partitioned by genes and codon position. In chapters III and IV, the program TIGER (Tree Independent Generation of Evolutionary Rates) (Cummins & McInerney 2011) was used to partition the data into character sets according to their rate of evolution. Each of the partitions can be assigned a suitable model. The effect of pruning the partition with the fastest evolving characters from the analysis was also explored to see whether this could help reducing "noise" and improve extracting the phylogenetic signal behind it.

Analyses were performed on desktop computers and the more computationally demanding ones in CIPRES (Miller *et al.* 2010), Bioportal (Kumar *et al.* 2009) and the cluster of the Laboratory of Genetics, Department of Biology, University of Turku.

3.5 Divergence time and diversification rate analyses

In Chapter II the divergence time estimation of the papilionoid lineages was

done with the program BEAST version 1.5.4. (Drummond & Rambaut 2007) with three fossils as calibration points. Patterns and changes in the tempo of diversification were analyzed with the software MEDUSA (Alfaro *et al.* 2009).

4. Results and Discussion

Chapter I: The enigmatic Lypusidae

The phylogenetic position of the genus *Lypusa* Zeller, 1982 was investigated using a large morphological dataset (Kaila, 2004) as a basis, including characters coded mostly from Gelechioidea (143 species), but also exemplar species representing the superfamilies Tineoidea, Gracillarioidea, Yponomeutoidea, Choreutoidea and Pyraloidea. In total, 188 morphological characters were coded from the adult male and female of *Lypusa tokari* (Elsner *et al.* 2008), but also from the larval case, and larval skin and pupal exuvia, thus providing for the first time detailed information on the morphology on the immature stages of *Lypusa*.

The result of the cladistic analysis did indeed refute a position of the genus with Tineoidea where it had previously been assigned to, yet never with a well-founded assignment to any particular family. Instead, *Lypusa* grouped within Gelechioidea, in the clade with species belonging to the family Amphisbatidae. Several morphological characters common to all species of the superfamily were found to be present in *Lypusa* providing support for a position in this superfamily, and not with the outgroup superfamilies. The most important of these characters are the scaled base of haustellum and pupal antennae mesially approaching each other thus invaginating the mesothoracic legs.

Morphological characters supporting a close affinity of *Lypusa* to *Pseudatemelia* and *Amphisbatis* include a larval case made of a folded piece of leaf, a densely porose larval head and a modification of the pupal abdominal segment 8.

Following the International Code of Nomenclature, the family name Lypusidae has priority over the younger family name Amphisbatidae, which has to yield in favor of the older name.

We also provide a summary of the information available on the biology of the genus.

Chapter II: Origin and diversification of the redefined butterflies

The results of the phylogenetic analyses generated by combined morphological data (45 larval, 32 pupal and 114 adult characters) and molecular data (eight gene regions) from 54 species representing most subfamilies of butterflies, skippers and butterfly moths, supported the paraphyly of the traditional conception of butterflies. In the traditional conception the families Papilionidae (swallow-tails), Pieridae (whites and sulphurs), Nymphalidae (brush-footed butterflies), the Lycaenidae (gossamer-winged

butterflies) and Riodinidae (metalmarks) belong to the same monopyletic group. However, our results support the position of Papilionidae as sister-group to the rest of the butterflies, Hesperiidae and Hedylidae. The results of the combined analysis also supported a strong sister-group relationship between Hedylidae and Hesperiidae. Two morphological characters, the flat projections on the mesophragma, and the shape of the third axillary sclerite at the base of the forewing, corroborate this result. The relationships between the rest of the butterflies were found to be (Pieridae + (Nymphalidae + (Riodinidae + Lycaenidae))), the position of Pieridae being, however, unstable in the analyses. Based on these results we suggest that the superfamily Papilionoidea be inclusive of Hesperiidae and Hedylidae.

The divergence time estimation indicated that the lineages leading to Papilionidae, Hesperiidae and Hedylidae and the rest of the butterflies diverged quite rapidly from each other in the Early Cretaceous, some 110 million years ago. Lineages leading to extant families had diverged from each other by 90 million years ago, with Pieridae diverging from the common ancestor at about 105 million years ago, and Nymphalidae from Lycaenidae and Riodinidae about 102 million years ago. Hedylids diverged from hesperiids about 99 million years ago and riodinids diverged from lycaenids about 88 million years ago. The results suggest that most within-family divergences leading to extant subfamily lineages occurred after the Cretaceous-Paleogene boundary (K–Pg boundary or also known as the K–T or Cretaceous–Tertiary boundary). This boundary is linked with the mass extinction of nearly three quarters of plant and animal species around 65 million years ago. Noteworthy patterns in the tempo of diversification is the slow speciation rate in the lineage leading to *Baronia brevicornis*, the sole species in the subfamily Baroniinae and considered a living fossil; a long delay before the species-rich skippers began to diversify; and a burst of diversification in the lineages leading to Nymphalidae, Riodinidae and Lycaenidae.

Chapter III: Revised classification for Gelechioidea

The phylogenetic analysis of the family relationships within the superfamily Gelechioidea is based on the densest taxon sampling (155 taxa) and largest morphological (139 adult, 49 pupal, and 65 larval characters) and molecular data (6127 base pairs) to date. The combination of morphological and molecular data improved support values for the groupings compared to analyses where either type of data were analyzed alone. Also, by combining data, we were able to avoid some of the unfavorable effects of the so-called rogue taxa in the analyses. Rogue taxa are unstable taxa which can take various positions in the trees thus lowering support values and affecting the resolution of the tree. Often such taxa are removed from the analyses to salvage the results concerning the remaining taxa (Wilkinson 1996; Thomson & Shaffer 2010; Trautwein *et al.* 2011). In our combined analyses several otherwise unstable taxa found a

stable position within the phylogenetic tree and we were able to include them.

The family level classification and delimitation of families of Gelechioidea (Hodges 1999; van Nieukerken *et al.* 2011) has been known to be in need of revision. Based on our results we propose a new classification and division of Gelechioidea into 16 families: Autostichidae, Lecithoceridae, Oecophoridae, Cosmopterigidae, Gelechiidae, Coleophoridae, Batrachedridae, Scythrididae, Blastobasidae, Stathmopodidae, Momphidae, Pterolonchidae, Depressariidae, Elachistidae, Xyloryctidae, and Lypusidae. The first eleven of these obtained strong support values in the model-based analyses, but the latter four rather weak support.

For the first time a monophyletic Oecophoridae s. s., (including Deuterogoniinae and Pleurotinae) was obtained with significant support. The entirely newly delineated and redefined Depressariidae Meyrick, 1883 includes the subfamilies Acriinae, Aeolanthinae, Cryptolechiinae, Depressariinae, Ethmiinae, Hypercalliinae, Hypertrophinae, Peleopodinae, Oditinae, Stenomatinae, *Carcina*, and a number of taxa without a former family position.

Elachistidae s. l. is found to be polyphyletic, and Elachistidae is restricted to include the subfamilies Agonoxeninae, Elachistinae, and Parametriotinae. Batrachedridae were found (as in Kaila *et al.* 2011) polyphyletic and restricted to the core *Batrachedra*. Other taxa previously included in this family grouped together with Coelopoetinae and Syringopainae in an expanded Pterolonchidae. Lypusidae s. s. and Chimabachidae formed a monophylum supported by the tongue-shaped, setose lobe on the male transtilla. Chimabachinae is newly united with Lypusidae as a subfamily.

The grouping of Stathmopodidae and Scythrididae is supported by a similarly expanded ductus seminalis. Both Stathmopodidae and Blastobasidae, which form a monophyletic lineage with Scythrididae, have a sclerotized ridge running from the lateral rod to the lateroposterior corners of tergum 1.

The results of the morphological examination confirm the paucity of morphological characters that can be used reliably to define families due to the high level of homoplasy. Apparently, species can sometimes only be placed to the correct family if both adult and immature stages are examined.

In general the deeper level relationships obtained very low support values and an even denser taxon sampling would certainly be beneficial. The place of the root, i.e. the direction of evolution within Gelechioidea, varied in the analyses, but was never within the monophyletic groups, which we use as the basis for our revised classification. The closest relatives of Gelechioidea remain unknown and a challenge for future research. Our taxon sampling was not designed to study subfamily relationships, but several of them appeared clearly to be in need of targeted studies.

Chapter IV: Morphological insight into the phylogeny of Ditrysia

The phylogeny of ditrysian Lepidoptera was studied by combining the hitherto largest morphological character dataset (530 larval, pupal, adult male and female characters coded from 318 species) with a dataset of eight gene regions from 422 taxa. The total number of exemplar taxa in the analyses was 473, which is the largest taxon sampling in studies on the phylogeny of the Ditrysia so far. The complete dataset was analyzed using maximum likelihood methods, the morphological dataset also with parsimony methods.

The resolution of the trees resulting from parsimony analyses was generally uniform, yet, interrelationships of superfamilies varied between the obtained most parsimonious trees. In addition, the adverse effect of several unstable taxa (e.g. *Euplocamus*, *Heliocosma* group, *Lactura*, Sematuridae, Mimallonidae, Hyblaeidae) effectively collapsed the consensus tree, even though the main structure of the obtained trees remained quite stable. The instability of several taxa was also seen in the low support values of several clades in the results of the maximum likelihood analyses. However, we opted for the inclusion of these taxa as their phylogenetic position is also of interest, and to be able to evaluate the evidence supporting the variable positions.

The combination of morphological data to molecular data in the analysis did not improve the resolution and low support values for the deeper nodes obtained in previous analyses based on Sanger-sequenced data. Our study confirms the paucity of morphological characters for defining larger assemblages of superfamilies. Also characters previously proposed to circumscribe subclades of Ditrysia were found to be to some extent equivocal and in some cases contradict recent results of phylogenetic studies based on molecular data. For example, the presence and absence of a character claimed to support the division between non-Apoditrysia and Apoditrysia, i.e. the lateral extensions of the abdominal sternum 2, was not clear-cut. The extension is absent in non-Apoditrysia, but was also found to be absent in several taxa obtaining a well-supported position within Apoditrysia in analyses based on genetic data.

Two characters previously considered to play key roles in the classification of Lepidoptera and used to define the clade Obtectomera; the presence of a setose lobe on the pulvilli and the obtect pupa, i.e. pupa with immobile intersegment between abdominal segments 3 and 4, were also checked across all superfamilies. The presence of the setose lobe on the pulvilli, was in many cases found to be ambiguous but also clearly absent in many taxa obtaining a position within Obtectomera in DNA-based analyses (e.g. Gelechioidea). Likewise, the mobility of pupal segments 3 and 4 was is several cases unclear and in many Yponomeutoidea the condition was found to be different from that of other Lepidoptera. The shape of the larval proleg, which has been used as a distinguishing feature to separate ‘Macrolepidoptera’

from 'Microlepidoptera', appeared to be convergent in several lineages. The division of Ditrysia to 'Macrolepidoptera' and 'Microlepidoptera' is not supported, at least in the traditional sense, by recent analyses based on genetic data. However, several interesting patterns emerged from combining morphological and molecular evidence, and new morphological evidence supporting the relatedness of several smaller groups was found.

Although the amount of morphological character data accounts only for 8% of the total amount of data in the combined analysis, the effect of morphological characters is strong. The positive effects of the inclusion of morphological data to the analyses was attested as several taxa with an unstable behavior in analyses based on DNA-data only found a stable position in the combined analyses, e.g. representatives of Cyclotornidae and Epipyropidae. Sesioidea were also found monophyletic in the combined analysis, when in molecules-only analyses they were not. In addition, when morphological and DNA data were in agreement, the support values for the clades were in general higher in the combined analyses than in those based on either type of data only.

However, combining data also caused some intriguing cases where morphology clearly overran the phylogenetic signal transmitted by the eight gene regions. For example, the position of Papilionoidea and the within-superfamily topology recovered in the present study, strongly contradict findings in several recent studies based on genetic data. Morphological characters causing the differing topology are most likely characters associated with a diurnal life mode e.g. clubbed antennae. Despite the conflicting signal, the support values were surprisingly high. Such patterns invite to explore in more detail the possibility and reasons of convergent evolution as an explanation of anatomical resemblance.

The main results of chapter IV based on the combined dataset include the recovery of a monophyletic Tineoidea, contradicting recent molecular studies. A new larval synapomorphy, a triangular cap dorsally covering the larval antenna, was found to unite Tineidae. *Metapherna salsa*, currently placed in Tineoidea, was repeatedly placed between Tineoidea and the Yponomeutoidea + Gracillarioidea clade in all the analyses. We suggest that *Metapherna* and allied taxa be assigned a family of their own.

The monophyly of Yponomeutoidea + Gracillarioidea was recovered supporting earlier DNA-based studies. The transverse costa, proposed as a possible synapomorphy of Yponomeutoidea, was also found in Roeslerstammiidae (Gracillarioidea).

Galacticoidea grouped with Tortricoidea in several analyses. Zygaenoidea, Sesioidea and Cossoidea are all monophyletic and in the same clade. Several small superfamilies (Carposinoidea, Millieriidae, *Tinagma*, *Tanaoctena*, Epermenioidea, Immoidea, Pterophoroidea, Alucitoidea and Choreutoidea) were left without a well-supported phylogenetic position in the larger context. However,

Schreckensteinioidea and Urodoidea share several larval synapomorphies and Immidae have features that indicate they could be related to Obtectomera. *Isonomeutis*, formerly considered a copromorphid, is in the clade with Alucitoidea. This placement is supported by larval synapomorphies that also suggest the paraphyly of Tineodidae as regards to Alucitidae.

Gelechioidea formed a monophylum, as did Pyraloidea, the division of which into Pyralidae and Crambidae is clear on morphological grounds. Hyblaeoidea (Hyblaeidea and Prodidactidae) were recovered as sister to Pyraloidea. Thyridoidea and Calliduloidea associate with Papilionoidea, the position of Thyridoidea being, however, somewhat unstable.

The Macroheterocera is one of the most strongly supported clades of the study and has Mimallonidae as sister-group. A monophyletic Geometroidea including Sematuridae + Epicopeiidae and Uranidae + Geometridae is obtained. *Doa* (Doidae) and *Axia* (Cimelioidea) are associated with Drepanoidea. Morphological evidence for these groupings is evaluated.

5. Conclusions and future directions

The focus of this thesis is on the examination of the morphology of ditrysian Lepidoptera in order to gain information on their evolutionary history and the relatedness of taxa (chapters I-IV). A central theme is also the importance of combining and comparing the phylogenetic signal obtained from morphology to that extracted from DNA-based data. As these studies show, both types of data carry valuable phylogenetic information and help to achieve a better understanding of the phylogeny and diversity of this enormous group of insects.

We studied the morphology of ditrysian Lepidoptera with the aim of resolving the phylogenetic position of single taxa but also that of larger assemblages. The articles of this thesis suggest that in ditrysian Lepidoptera phylogenetically informative morphological characters are abundant at the species, family and superfamily level, but characters that could be used to infer phylogenetic relationships between superfamilies or larger assemblages are scarce. Similar results have been observed in studies based on DNA-based data. The deeper evolutionary relationships between the major ditrysian lineages are thus left as a challenge for future research, both genetic and morphological.

The results also show that there are several benefits from combining and comparing morphological data. When the phylogenetic signal emitted from both morphological and molecular data is in agreement, support for, and therefore confidence in the proposed relationships, is of course stronger. On the other hand, when the signal of morphological and molecular data is conflicting, it encourages looking for causes in the methods used, but it is also an opportunity to learn about moths

and butterflies from the data. In evaluating conflicting results we can seek for explanations in the underlying biological mechanisms causing this conflict, e.g. by identifying cases of convergent evolution and differences in rates of evolution. Combining morphological and molecular data also allows securing a place in the phylogenetic tree for some taxa that are unstable in analyses in which only either type of data is used.

The morphological examination in the studies of this thesis was done with light microscopes and concentrated on external and sclerotized characters. However, valuable data can also be coded from other structures and future studies should focus on those. Research in comparative morphology on Lepidoptera has very long traditions and there is an extremely rich literature with detailed illustrations on characters not included in the studies of this thesis, e.g. muscles and glands. These observations should also be compared across larger groups of Lepidoptera, and subjected to a phylogenetic analysis, as they most certainly carry phylogenetic information. Also, acknowledging the usefulness of morphological information in phylogenetic systematics, it is clear that research on Ditrysia could only gain from the advances in high throughput digital imaging and non-invasive imaging techniques. In addition, modern methods to visualize and share morphological information make it more accessible and reduce concerns about the subjectivity of observations. Although comparative morphology has a reputation for being time-consuming, it is a learning process in which skills constantly develop, making the process faster. The possibility to make direct observations, and often find signs revealing something of the evolutionary history of the organism makes the work even more satisfying.

The morphological datasets of this thesis are among the largest collected on ditrysian groups and will hopefully be useful in future research on Ditrysia as a whole or in targeted studies on the two superfamilies also focused on in this thesis, Gelechioidea and Papilionoidea. We aimed to collect characters which could be used in revised identification keys. Knowledge of morphological features characteristic of diverse groups of Ditrysia may also allow identification of taxa from which sequence data no longer can be extracted e.g. museum samples and fossils. A benefit of incorporating morphological data to phylogenetic analyses is that it enables including fossils as terminal taxa.

As the evolutionary tree for Lepidoptera becomes more robust, the next step will be to obtain estimates on the times of divergence of the major lineages. This is also a place for morphological data to show its importance and usefulness. Dated fossils which can be securely assigned a position in the phylogenetic tree can be used as calibration points in estimating the times of divergence of lineages. Divergence time estimates allow us to explore in more depth the evolutionary history of Ditrysia, and learn when and why certain morphological or behavioral traits have arisen, how they have evolved, and can they possibly explain the success of certain lineages.

Acknowledgements

The motivating topic, combined with excellent guidance, were the two most important factors that enabled me to carry out this work. Therefore, my deepest gratitude goes to my supervisor Lauri Kaila. I am most grateful for having been offered such a fascinating research topic. He tirelessly found time to help me with my process. I have enjoyed the years of our collaboration and have learned a lot from him. I also wish to thank Lauri for his patience and understanding of the fact that I often needed to prioritize my family during this process.

Completing the thesis would not have been possible without the support of the great research group I had the pleasure of being part of. I am thankful to Marko Mutanen, co-author of three articles in the thesis, for his help in molecular methods and manuscript preparation. Niklas Wahlberg, whom I know to be absolutely swamped with work, but despite that, has always had the time to help me with the analyses. Not once has he left my emails unanswered. Mari Kekkonen provided excellent peer support, which was needed particularly during the last weeks of this project. I have also tremendously enjoyed our many memorable conference trips abroad. Pasi Sihvonen is acknowledged for his help in writing the successful academy proposal that launched the project this thesis is part of. He has also helped with material preparation and commented on manuscripts. Carlos Peña has been most helpful and patient in answering questions related to the fantastic Voseq database.

Many people outside the research group have also helped during the research process and in preparing articles ready for publication. I wish to express my gratitude to the library assistant of the museum, Marjatta Mikkonen, who will go through granite and grey stone to find the publications needed. Museum assistants Elvira Rättel and Ritva Talman are wizards in sample preparation. I thank Pekka Malinen for providing help with the cameras and related software, and he has kindly let me use his fantastic photographs in many of my presentations. Marianna Teräväinen and Katja&Nathan Thomson have checked the language of several of my manuscripts and other texts. Kaisaleena Halinen is the artist behind the picture on the cover of this thesis. Heidi Viljanen, Sanna Saari and Ville Heimala have saved me from having to struggle with layout and graphics editing programs. I would like to thank Jyrki Muona and Jaakko Hyvönen for explaining many difficult aspects of systematics. They are also acknowledged for organizing discussion groups and for their attempts to keep the level of knowledge of systematics of both PhD and MSc students at a high level. I am grateful to Olof Biström for writing several support letters to the Finnish-Swedish foundations I ended up receiving funding from. The research visit to the Smithsonian Institution would not have been as successful as it was without the help of our excellent host, Dr John Brown. I will never forget the hospitality John and his lovely wife, Poody, showed during our visit. I also feel that several people I encountered at the meetings of Societas Europaea Lepidopterologica and Lepidopterists'

Society have helped me enter and become part of the international community of lepidopterists.

The museum has an active group of PhD students that meets regularly. The atmosphere in these meetings has been open and the peer support sincere. I feel that I have greatly benefited from the encouragement and good advice shared during these meetings. Thank you, Aino, Andrea, Diego, Elina, Hanna La., Hanna Li., Heidi, Juho, Mari, Maria, Marianna, Sanna and Sara! Doing a PhD at the museum would not have been the same without you!

The zoology unit has an enjoyable and motivating working atmosphere, which has certainly supported the progress of my work. I wish to thank the director of the museum, Leif Schulman, and the unit director, Aino Juslén, for making sure that the PhD students feel part of this community and that their problems are heard. The wonderful colleagues have been a great reason to come to work. I have learned something new about biodiversity from them every day (although, I am also happy that my colleagues can talk about other topics, too!). I wish to thank my lunch company: Katja Nylund, Anne Koivunen, Marianna Teräväinen, Sanna Saari, Heidi Viljanen, Ville Heimala and Silva Sallamaa. Particularly interesting conversations took place at the KeKe (keskiviikkokerho/Wednesday club) lunches.

Extra time, money and moral support were often needed during my PhD process. I want to express my gratitude to my parents-in-law, Arja and Jussi Heikkilä. Their commitment to taking care of our sons has been without limits. They have regularly looked after them allowing me to work longer days. During this time the grandparents defied all kinds of flu viruses and carried on even when the boys were sick. It has also been absolutely wonderful to come home to a clean house after a long day at a messy desk.

I received invaluable encouragement and support from both my sister and parents. They are the people with whom I can show my stress and anxiety, and they are the ones who can also alleviate it. I want to thank my sister Silja, my mother Laura, who to our great sorrow left us too soon, and my father Tuomas for their interest in my work and their wise advice. My brother-in-law, Sakari, has also kindly sat through the practice rounds of my first presentations to be held at big international conferences!

I owe my greatest thanks to my husband Tomi and our two sons, Eero and Tuomo. The past years have been very busy at home, but it is this house full of life that has given me the energy to work and the courage to take on new challenges. Thank you, Tomi, for taking such good care of the kids (and me). Especially during the last months running the whole show has been completely on your shoulders. From now on, I will try to come home by five. I hope dinner is ready! :)

The following foundations have supported the work: Academy of Finland, Finnish Cultural Foundation, Ella and Georg

Ehrnrooth Foundation, Oscar Öflund Foundation and the University of Helsinki. Attendance in conferences, workshops and scientific visits abroad were made possible by grants received from Societas Entomologica Fennica, Societas Entomologica Helsingforsiensis, the Chancellor's Travel Grant (University of Helsinki), the Willi Hennig Society and LUOVA (Doctoral Programme in Wildlife Biology Research, University of Helsinki).

References

Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G. & Harmon, L. (2009). Nine exceptional radiations plus high turnover explain species diversity of jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 13410-13414.

Barth, F.G. (1985). *Insects and flowers. The biology of a partnership*. George Allen & Unwin, London.

Bazinet, A.L., Cummings, M.P., Mitter, K.T. & Mitter, C.W. (2013). Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An exploratory study. *PLoS ONE*, 8, e82615.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, 21, 163-193.

Boussau, B., Walton, Z., Delgado, J.A., Collantes, F., Beani, L., Stewart, I.J., Cameron, S.A., Whitfield, J.B., Johnston, J.S., Holland, P.W.H., Bachtrog, D., Kathirithamby, J. & Huelsenbeck, J.P. (2014). Strepsiptera, phylogenomics and the long branch attraction problem. *PLoS ONE*, 9, e107709.

Bucheli, S.R. & Wenzel, J. (2005). Gelechioidea (Insecta : Lepidoptera) systematics: A reexamination using combined morphology and mitochondrial DNA data. *Molecular Phylogenetics and Evolution*, 35, 380-394.

Chiu, L.C.M., Kong, C.K.L. & Ooi, V.E.C. (2003). Antiproliferative effect of chlorophyllin derived from a traditional Chinese medicine Bombyx mori excreta on human breast cancer MCF-7 cells. *International Journal of Oncology*, 23, 729-735.

Cho, S., Zwick, A., Regier, J.C., Mitter, C., Cummings, M.P., Yao, J., Du, Z., Zhao, H., Kawahara, A.Y., Weller, S., Davis, D.R., Baixeras, J., Brown, J.W. & Parr, C. (2011). Can deliberately incomplete gene sample augmentation improve a phylogeny estimate for the advanced moths and butterflies (Hexapoda: Lepidoptera)? *Systematic Biology*, 60, 782-796.

Cummins, C.A. & McInerney, J.O. (2011). A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Systematic Biology*, 60, 833-844.

de Carvalho, M., Bockmann, F., Amorim, D. & Brandão, C. (2008). Systematics must embrace comparative biology and evolution, not speed and automation. *Evolutionary Biology*, 35, 150-157.

DeFoliart, G.R. (1989). The human use of insects as food and as animal feed. *Bulletin of the Entomological Society of America*, 35, 22-36.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4, 699–710.

Drummond, A.J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, 214.

Dugdale, J.S. (1974). Female genital configuration in the classification of Lepidoptera. *New Zealand Journal of Zoology,* 14, 375-383.

Elsner, G., Liska, J. & Petru, M. (2008). Eine neue Art der Gattung Lypusa. *Entomologische Zeitschrift*, 118, 107-112.

Emmet, A.M. (1991). *The scientific names of the British Lepidoptera. Their history and meaning*. Harley Books, Colchester.

Forest, F. (2009). Calibrating the Tree of Life: fossils, molecules and evolutionary timescales. *Annals of Botany*, 104, 789–794.

Fox, R., Randle, Z., Hill, L., Anders, S., Wiffen, L. & Parsons, M.S. (2011). Moths count: Recording moths for conservation in the UK. *Journal of Insect Conservation*, 15, 55-68.

Gatesy, J. (2005). Hidden likelihood support in genomic data: can forty-five wrongs make a right? *Systematic Biology*, 54, 483-492.

Giribet, G. (2010). A new dimension in combining data? The use of morphology and phylogenomic data in metazoan systematics. *Acta Zoologica*, 91, 11-19.

Goloboff, P.A. (1993). NONA Version 2. In. Published by the author Tucumán.

Goloboff, P.A., Farris, J.S. & Nixon, K.C. (2000). TNT (Tree Analysis using New Technology)

(BETA). In. Published by the authors Tucumán.
Goloboff, P.A., Farris, J.S. & Nixon, K.C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics*, 24, 774-786.
Grimaldi, D. (1999). The co-radiations of pollinating insects and angiosperms in the Cretaceous. *Annals of the Missouri Botanical Garden*, 86, 373-406.
Grimaldi, D. & Engel, M.S. (2005a). Amphiesmenoptera: The Caddisflies and Lepidoptera. In: *Evolution of the Insects*. Cambridge University Press Cambridge, pp. 548-606.
Grimaldi, D. & Engel, M.S. (2005b). Diversity and Evolution. In: *Evolution of the Insects*. Cambridge University Press Cambridge, pp. 1-41.
Heikkilä, M., Kaila, L., Mutanen, M., Peña, C. & Wahlberg, N. (2012). Cretaceous origin and repeated tertiary diversification of the redefined butterflies. *Proceedings of the Royal Society B*, 279: 1731, 1093-1099.
Heikkilä, M., Mutanen, M., Kekkonen, M. & Kaila, L. (2014). Morphology reinforces proposed molecular phylogenetic affinities: a revised classification for Gelechioidea (Lepidoptera). *Cladistics* 30, 563-589.
Hermsen, E.J. & Hendricks, J.R. (2008). W(h)ither fossils? Studying morphological character evolution in the age of molecular sequences. *Annals of the Missouri Botanical Garden*, 95, 72-100.
Hodges, R.W. (1999). The Gelechioidea. In: *Handbook of Zoology. Lepidoptera, Moths and Butterflies. Vol. 1: Evolution, Systematics and Biogeography* (ed. Kristensen NP). Walter de Gruyter Berlin, New York, pp. 131-1158.
Holloway, J.D. & Nielsen, E.S. (1999). Biogeography of the Lepidoptera. In: *Handbook of Zoology. Lepidoptera, Moths and Butterflies. Vol. 1: Evolution, Systematics and Biogeography* (ed. Kristensen NP). Walter de Gruyter Berlin, New York, pp. 423-462.
Huelsenbeck, J.P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754-755.
Kaila, L. (2004). Phylogeny of the superfamily Gelechioidea (Lepidoptera: Ditrysia): an exemplar approach. *Cladistics*, 20, 303-340.
Kaila, L., Mutanen, M. & Nyman, T. (2011). Phylogeny of the mega-diverse Gelechioidea (Lepidoptera): Adaptations and determinants of success. *Molecular Phylogenetics and Evolution*, 61, 801-809.
Kawahara, A., Regier, J., Davis, D., Kawakita, A., Mitter, C., Ohshima, I., De Prins, J., Lopez-Vaamonde, C. & Wagner, D. (2008). Molecular phylogeny of the Gracillarioidea: Preliminary evidence from eight genes. *56th Annual Meeting of the Entomological Society of America (ESA 2008)*.
Kawahara, A.Y. & Breinholt, J.W. (2014). Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proceedings of the Royal Society B*, 281, 1788.
Kikuchi, H., Takahashi, N. & Oshima, Y. (2004). Novel aromatics bearing 4-O-methylglucose unit isolated from the oriental crude drug *Bombyx batryticatus*. *Tetrahedron Letters: International Organ for the Rapid Publication of Preliminary Communications in Organic Chemistry*, 45, 367.
Kitching, I.J. & Rawlins, J.E. (1999). The Noctuoidea. In: *Handbook of Zoology. Lepidoptera, Moths and Butterflies. vol 1: Evolution, Systematics and Biogeography* (ed. Kristensen NP). Walter de Gruyter Berlin, pp. 355-401.
Kolaczkowski, B. (2009). Long-branch attraction bias and inconsistency in Bayesian phylogenetics. *PLoS ONE*, 4 e7891.
Kristensen, N.P. (2011). Lepidoptera systematics 2011: state and challenges. *XVIIth European Congress of Lepidopterology. Congress Abstracts*.
Kristensen, N.P., Scoble, M.J. & Karsholt, O. (2007). Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity. *Zootaxa*, 1668, 699-747.
Kristensen, N.P. & Skalski, A.W. (1999). Phylogeny and palaeontology. In: *Handbook of Zoology IV: Lepidoptera, moths and butterflies. Vol.1: Evolution, systematics and biogeography* (ed. Kristensen NP). Walter de Gruyter Berlin, New York, pp. 7-25.
Kumar, S., Skjaeveland, A., Orr, R.J.S., Enger, P., Ruden, T., Mevik, B.H., Burki, F., Botnen, A. & Shalchian-Tabrizi, K. (2009). AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *Bmc Bioinformatics*, 10, 357.
Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29, 1695-1701.
Lewis, P.O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50, 913-925.
Lopez-Vaamonde, C., Kawahara, A., Rougerie, R., Ohshima, I., Kawakita, A., Bouteleux, O., De Prins, J. & Lees, D.C. (2014). DNA barcod-

ing reveals that the reverse latitudinal gradient of Gracillariidae is an artefact of tropical undersampling. In: *Forum Herbulot 2014; "How to accelerate the inventory of Biodiversity"* Schlettau (Germany).

Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop, GCE 2010*.

Minet, J. (1986). Ebauche d'une classification moderne de l'ordre des lépidoptères. *Alexanor*, 14, 291-313.

Minet, J. (1991). Tentative reconstruction of the ditrysian phylogeny (Lepidoptera, Glossata). *Entomologica Scandinavica*, 22, 69-95.

Mutanen, M., Wahlberg, N. & Kaila, L. (2010). Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proceedings of the Royal Society B*, 277, 2839-2848.

Niehuis, O., Naumann, C.M. & Misof, B. (2006). Phylogenetic analysis of Zygaenoidea small-subunit rRNA structural variation implies initial oligophagy on cyanogenic host plants in larvae of the moth genus Zygaena (Insecta : Lepidoptera). *Zoological Journal of the Linnean Society*, 147, 367-381.

Nixon, K.C. (2002). Winclada ver. 1.00.08. *Published by the author, Ithaca, NY.* Available at http://www.cladistics.com.

Nixon, K.C. & Carpenter, J.C. (1996). On simultaneous analysis. *Cladistics*, 12, 221-241.

O'Hara, R.J. (1988). Homage to Clio, or, toward an historical philosophy for evolutionary biology. *Systematic zoology*, 37, 142-155.

O'Leary, M.A. & Kaufman, S.G. (2012). MorphoBank 3.0: Web application for morphological phylogenetics and taxonomy. http://www.morphobank.org

Peña, C. & Malm, T. (2012). VoSeq: a voucher and DNA sequence web application. *PLoS ONE*, 7, e39071.

Powell, J.A., Mitter, C. & Farrell, B. (1999). Evolution of larval food preferences in Lepidoptera. In: *Handbook of Zoology. Lepidoptera, Moths and Butterflies. vol 1: Evolution, Systematics and Biogeography* (ed. Kristensen NP). Walter de Gruyter Berlin, New York, pp. 403-422.

Ramos-Elorduy, J., Moreno, J.M.P., Vazquez, A.I., Landero, I., Oliva-Rivera, H. & Camacho, V.H.M. (2011). Edible Lepidoptera in Mexico: Geographic distribution, ethnicity, economic and nutritional importance for rural people. *Journal of Ethnobiology & Ethnomedicine*, 7, 2-23.

Regier, J.C., Cook, C.P., Mitter, C. & Hussey, A. (2008). A phylogenetic study of the 'bombycoid complex' (Lepidoptera) using five protein-coding nuclear genes, with comments on the problem of macrolepidopteran phylogeny. *Systematic Entomology*, 33, 175-189.

Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.-C., Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C. & Mitter, K.T. (2013). A Large-Scale, Higher-Level, Molecular Phylogenetic Study of the Insect Order Lepidoptera (Moths and Butterflies). *PLoS ONE*, 8, e58568.

Regier, J.C., Zwick, A., Cummings, M.P., Kawahara, A.Y., Cho, S., Weller, S., Roe, A., Baixeras, J., Brown, J.W., Parr, C., Davis, D.R., Epstein, M., Hallwachs, W., Hausmann, A., Janzen, D.H., Kitching, I.J., Solis, M.A., Yen, S.H., Bazinet, A.L. & Mitter, C. (2009). Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): An initial molecular study. *BMC Evolutionary Biology*, 9, 280.

Regier, M., Mitter, C., Davis, D.R., Harrison, T.L., Sohn, J.-E., Cummings, M.P., Zwick, A. & Mitter, K.T. (2014). A molecular phylogeny and revised classification for the oldest ditrysian moth lineages (Lepidoptera: Tineoidea), with implications for ancestral feeding habits of the mega-diverse Ditrysia. *Systematic Entomology*.

Roe, A.D., Weller, S.J., Baixeras, J., Brown, J., Cummings, M.P., Davis, D.R., Kawahara, A.Y., Parr, C., Regier, J.C., Rubinoff, D. & Simonsen, T.J. (2010). Evolutionary framework for Lepidoptera model systems. In: *Genetics and molecular biology of Lepidoptera* (eds. Goldsmith M & Marec F). CRC Press Boca Raton, FL, pp. 1-24.

Rokas, A. & Carroll, S.B. (2006). Bushes in the tree of life. *PLoS Biology*, 4, e352.

Rokas, A., Williams, B.L., King, N. & Carroll, S.B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, 425, 798.

Rolland, J., Cadotte, M.W., Davies, J., Devictor, V., Lavergne, S., Mouquet, N., Pavoine, S., Rodrigues, A., Thuiller, W., Turcati, L., Winter, M., Zupan, L., Jabot, F. & Morlon, H. (2012). Using phylogenies in conservation: new perspectives. *Biology Letters*, 8, 692-694.

Scoble, M.J. (1992). *The Lepidoptera : form, function, and diversity*. Natural History Museum in association with Oxford University Press, Ox-

ford.
Scudder, G.G.E. (2009). The importance of insects. In: *Insect biodiversity: science and society* (eds. Foottit RG & Adler PH). Wiley-Blackwell Oxford, pp. 7-32.
Solis, M.A. (2007). Phylogenetic studies and modern classification of the Pyraloidea (Lepidoptera). *Revista Colombiana De Entomologia*, 33, 1-9.
Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22, 2688-2690.
Talavera, G. & Vila, R. (2011). What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evolutionary Biology*, 11, 315.
Talekar, N.S. (1993). Biology, ecology, and management of the Diamondback moth. *Annual Review of Entomology*, 38, 275-301.
Tao, N., Richardson, R., Bruno, W. & Kuiken, C. (2010). FINDMODEL. In: Los Alamos, NM: Los Alamos National Security. See http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html.
Thomson, R.C. & Shaffer, H.B. (2010). Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Systematic Biology*, 59, 42-58.
Timmermans, M.J.T.N., Lees, D.C. & Simonsen, T.J. (2014). Towards a mitogenomic phylogeny of Lepidoptera. *Molecular Phylogenetics and Evolution*, doi: 10.1016/j.ympev.2014.05.031.
Trautwein, M.D., Wiegmann, B.M. & Yeates, D.K. (2011). Overcoming the effects of rogue taxa: Evolutionary relationships of the bee flies. *PLoS currents*, 3, RRN1233-RRN1233.
Wahlberg, N. & Wheat, C.W. (2008). Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology*, 57, 231-242.
Wahlberg, N., Wheat, C.W. & Pena, C. (2013). Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS ONE*, 8.
van Nieukerken, E.J., Kaila, L., Kitching, I.J., Kristensen, N.P., Lees, D.C., Minet, J., Mitter, C., Mutanen, M., Regier, J.C., Simonsen, T.J., Wahlberg, N., Yen, S.H., Zahiri, R., Adamski, D., Baixeras, J., Bartsch, D., Bengtsson, B.Å., Brown, J.W., Bucheli, S.R., Davis, D.R., De Prins, J., De Prins, W., Epstein, M.E., Gentili-Poole, P., Gielis, C., Hättenschwiler, P., Hausmann, A., Holloway, J.D., Kallies, A., Karsholt, O., Kawahara, A., Koster, S.J.C., Kozlov, M., Lafontaine, J.D., Lamas, G., Landry, J.F., Lee, S., Nuss, M., Penz, C., Rota, J., Schmidt, B.C., Schintlmeister, A., Sohn, J.C., Solis, M.A., Tarmann, G.M., Warren, A.D., Weller, S., Yakovlev, R., Zolotuhin, V. & Zwick, A. (2011). Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa*, 3148, 212-221.
Warren, A.D., Ogawa, J.R. & Brower, A.V.Z. (2009). Revised classification of the family Hesperiidae (Lepidoptera: Hesperioidea) based on combined molecular and morphological data. *Systematic Entomology*, 34, 467-523.
Wenzel, J.W. (2002). Phylogenetic analysis: the basic method. In: *Techniques in molecular systematics and evolution* (eds. DeSalle R, Giribet G & Wheeler W). Birkhaeuser Verlag Basel, Boston & Berlin, pp. 4-30.
Whalley, P.E.S. (1985). The systematics and palaeogeography of the Lower Jurassic insects of Dorset, England. *Bulletin of the British Museum (Natural History) Geology*, 39, 107-189.
Wheeler, Q.D. (2008). Undisciplined thinking: morphology and Hennig's unfinished revolution. *Systematic Entomology*, 33, 2-7.
Whitfield, J.B. & Kjer, K.M. (2008). Ancient rapid radiations of insects: Challenges for phylogenetic analysis. *Annual Review of Entomology*, 53, 449-472.
Whitfield, J.B. & Lockhart, P.J. (2007). Deciphering ancient rapid radiations. *Trends in Ecology & Evolution*, 22, 258-265.
Wiens, J.J. (2004). The role of morphological data in phylogeny reconstruction. *Systematic Biology*, 53, 653-661.
Wilkinson, M. (1996). Majority-rule reduced consensus trees and their use in bootstrapping. *Molecular Biology and Evolution*, 13, 437.
Winfree, R., Gross, B.J. & Kremen, C. (2011). Valuing pollination services to agriculture. *Ecological Economics*, 71, 80-88.
Wright, A.M. & Hillis, A.M. (2014). Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE*, 9 e109210.
Zahiri, R., Kitching, I.J., Lafontaine, J.D., Mutanen, M., Kaila, L., Holloway, J.D. & Wahlberg, N. (2011). A new molecular phylogeny offers hope for a stable family level classification of the Noctuoidea (Lepidoptera). *Zoologica Scripta*, 40, 158-173.